

Experimental boson sampling

Justin B. Spring^{1*}, Benjamin J. Metcalf¹, Peter C. Humphreys¹,
W. Steven Kolthammer¹, Xian-Min Jin^{1,2}, Marco Barbieri¹, Animesh Datta¹,
Nicholas Thomas-Peter¹, Nathan K. Langford^{1,3}, Dmytro Kundys⁴,
James C. Gates⁴, Brian J. Smith¹, Peter G.R. Smith⁴, and Ian A. Walmsley^{1*}

¹Clarendon Laboratory, Department of Physics, University of Oxford, OX1 3PU, UK

²Department of Physics, Shanghai Jiao Tong University, Shanghai 200240, PR China

³Department of Physics, Royal Holloway, University of London, TW20 0EX, UK

⁴Optoelectronics Research Centre, University of Southampton, Southampton, SO17 1BJ, UK

*To whom correspondence should be addressed

E-mail: j.spring1@physics.ox.ac.uk, i.walmsley1@physics.ox.ac.uk

While universal quantum computers ideally solve problems such as factoring integers exponentially more efficiently than classical machines, the formidable challenges in building such devices motivate the demonstration of simpler, problem-specific algorithms that still promise a quantum speedup. We construct a quantum boson sampling machine (QBSM) to sample the output distribution resulting from the nonclassical interference of photons in an integrated photonic circuit, a problem thought to be exponentially hard to solve classically. Unlike universal quantum computation, boson sampling merely requires indistinguishable photons, linear state evolution, and detectors. We benchmark our QBSM with three and four photons and analyze sources of sampling inaccuracy. Our studies pave the way to larger devices that could offer the first definitive quantum-enhanced computation.

Universal quantum computers require physical systems that are well-isolated from the decohering effects of their environment, while at the same time allowing precise manipulation during computation. They also require qubit-specific state initialization, measurement, and the generation of quantum correlations across the system (1–5). Although there has been substantial progress in proof-of-principle demonstrations of quantum computation (6–10), si-

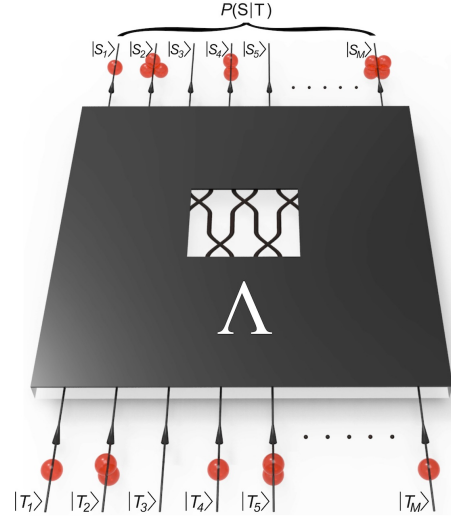


Figure 1: Model of quantum boson sampling. Given a specified initial number state $|\mathbf{T}\rangle = |T_1 \dots T_M\rangle$ and linear transformation Λ , a quantum boson sampling machine efficiently samples from the distribution $P(\mathbf{S}|\mathbf{T})$ of possible outcomes $|\mathbf{S}\rangle = |S_1 \dots S_M\rangle$.

multaneously meeting these demands has proven difficult. This motivates the search for schemes that can demonstrate quantum-enhanced computation under more favorable experimental conditions. The power of one qubit (11), permutational (12) and instantaneous (13) quantum computation, are examples that have allowed investigation of the space be-

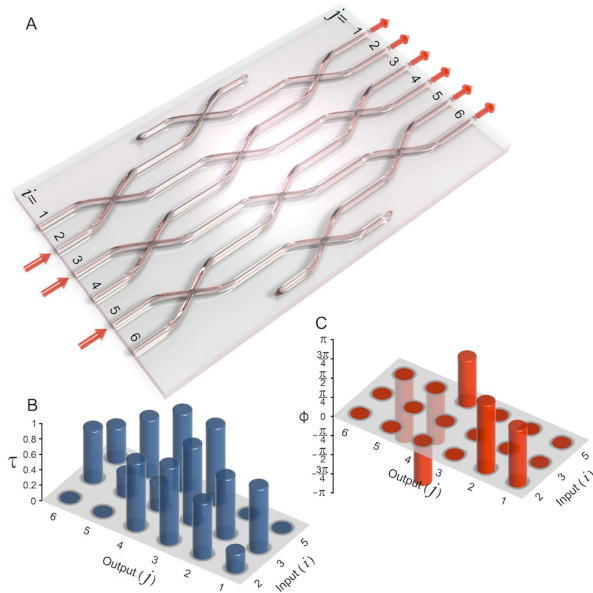


Figure 2: Schematic and characterization of the photonic circuit. (a) The silica-on-silicon waveguide circuits consist of $M = 6$ accessible spatial modes (labeled 1–6). For the three-photon experiment, we launch single photons into inputs $i = 2, 3$ and 5 from two parametric down-conversion sources and postselect outcomes in which three detections are registered amongst the output modes j . For the four-photon experiment, which is implemented on a different chip of identical topology, we inject a double photon pair from a single source into the modes $i = 1, 3$ and postselect on four detection events. (b)-(c) Measured elements of the linear transformation $\Lambda_{ij} = \tau_{ij} e^{i\phi_{ij}}$ linking the input mode i to the output mode j of our three-photon apparatus. The circuit topology dictates that several τ_{ij} are zero, and our phase-insensitive input states and detection methods imply only six non-zero ϕ_{ij} . Since only relative values are needed due to post-selection, we rescale each row of τ so that its maximum value is unity.

tween classical and universal quantum machines.

Boson sampling has recently been proposed as a specific quantum computation that is more efficient than its classical counterpart but only requires identical bosons, linear evolution, and measurement (14). The distribution of bosons that have passed through a linear system represented by the transformation Λ is thought to be exponentially hard to sample from classically (14). The probability amplitude of obtaining a certain output is directly proportional to the permanent of a corresponding submatrix of Λ . The permanent expresses the wavefunction of identical bosons, which are symmetric under exchange (15, 16); in contrast, the Slater determinant expresses the

wavefunction of identical fermions, which are antisymmetric under exchange. While determinants can be evaluated efficiently, permanents have long been believed to be hard to compute (17); the best known algorithm scales exponentially with the size of the matrix (18). One can envision a race between a classical and a quantum machine to sample the boson distribution given an input state and Λ . The classical machine would evaluate at least part of the probability distribution, which requires the computation of matrix permanents. The QBSM instead creates indistinguishable bosons, directly implements Λ , and records the outputs. For a sufficiently large problem size, the quantum machine is expected to win (14).

Photonics is a natural platform to implement boson sampling since sources of indistinguishable photons are well-developed (19), and integrated optics offers a scalable route to large linear systems (20). In addition, photons have extremely low decoherence rates. Importantly, boson sampling requires neither nonlinearities nor on-demand entanglement, unlike photonic approaches to universal quantum computation (21). This clears the way for experimental boson sampling with existing photonic technology, building on the extensively studied two-photon Hong-Ou-Mandel (HOM) interference effect (22).

A QBSM (Fig. 1) samples the output distribution of a multi-particle bosonic quantum state $|\Psi_{\text{out}}\rangle$, prepared from a specified initial state $|\mathbf{T}\rangle$ and linear transformation Λ . A trial begins with the input state $|\mathbf{T}\rangle = |T_1 \dots T_M\rangle \propto \prod_{i=1}^M (\hat{a}_i^\dagger)^{T_i} |0\rangle$, which describes $N = \sum_{i=1}^M T_i$ particles distributed in M input modes in the occupation-number representation. The output state $|\Psi_{\text{out}}\rangle$ is generated according to the linear map from output to input mode creation operators $\hat{a}_i^\dagger = \sum_{j=1}^M \Lambda_{ij} \hat{b}_j^\dagger$, where Λ is an $M \times M$ matrix. Finally, the particles in each of the M output modes are counted. The probability of a particular measurement outcome $|\mathbf{S}\rangle = |S_1 \dots S_M\rangle$ is given by

$$P(\mathbf{S}|\mathbf{T}) = |\langle \mathbf{S} | \Psi_{\text{out}} \rangle|^2 = \frac{|\text{Per}(\Lambda^{(\mathbf{S}, \mathbf{T})})|^2}{\prod_{j=1}^M S_j! \prod_{i=1}^M T_i!} \quad (1)$$

where the $N \times N$ submatrix $\Lambda^{(\mathbf{S}, \mathbf{T})}$ is obtained by keeping S_j (T_i) copies of the j^{th} column (i^{th} row)

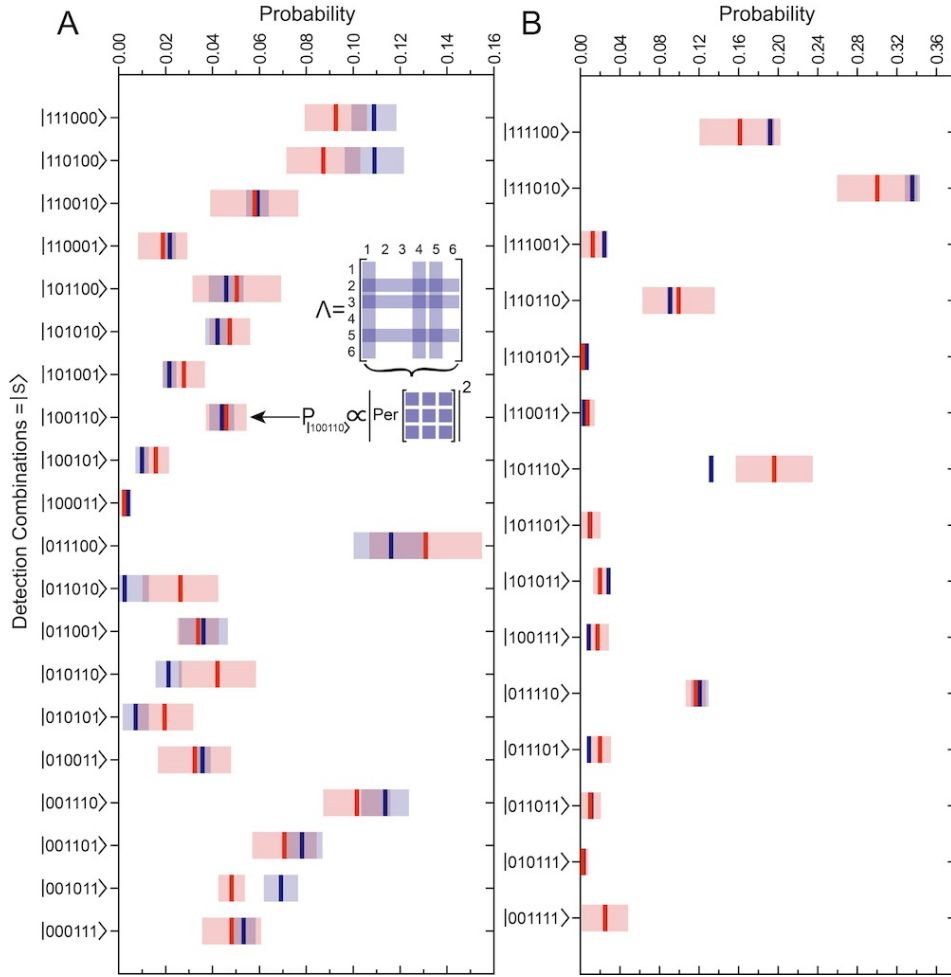


Figure 3: Boson sampling results. The measured relative frequencies \mathbf{P}^{exp} of outcomes in which the photons are detected in distinct modes are shown in red for (a) three- and (b) four-photon experiments. Each data set is collected over 160 hours, and statistical variations in counts are shown by the red shaded bars. The theoretical distributions \mathbf{P}^{th} (blue) are obtained from the permanent of a submatrix constructed from the full transformation Λ , as depicted in the inset. The blue error bars arise from uncertainties in the characterization of Λ .

of Λ (23).

The aim of boson sampling is to sample from the distribution given by $P(\mathbf{S}|\mathbf{T})$. A QBSM achieves this efficiently by directly sampling $|\langle \mathbf{S} | \Psi_{\text{out}} \rangle|^2$. The classical strategy, on the other hand, requires an estimation of the right-hand side of Eqn. (1), for which the calculation time scales exponentially in N (18).

Our QBSM consists of sources of indistinguishable single photons, a multiport linear optical circuit, and single-photon counting detectors. Two parametric down-conversion (PDC) pair sources (24) are used to inject up to four photons into a silica-on-silicon integrated photonic circuit, fabricated by UV writing (25). The circuit is shown in

Fig. 2a and consists of $M=6$ input and output spatial modes coupled by a network of ten beam splitters (20). The output state is measured with single-photon avalanche photodiodes on each mode. We only consider outcomes in which the number of detections equals the intended number of input photons (23).

Our central result of three- and four-boson sampling is shown in Fig. 3. In the first case, we repeatedly inject three photons in the input state $|\mathbf{T}\rangle = |011010\rangle$, monitor all outputs, and collect all three-fold coincident events. In the four-photon experiment, we use the input $|\mathbf{T}\rangle = |202000\rangle$ and record

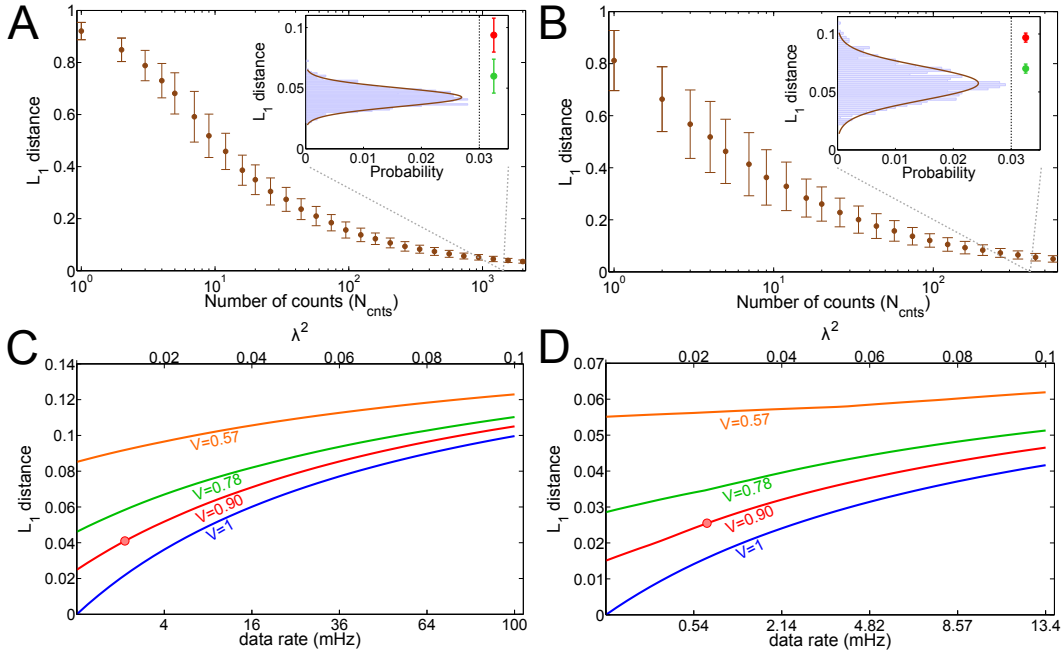


Figure 4: Sampling accuracy. We consider several boson outcome distributions: the experimental samples \mathbf{P}^{exp} , the ideal predictions of Eqn. (1) \mathbf{P}^{th} , and the predictions of the full model \mathbf{P}^{mod} . (a) The L_1 distance d between \mathbf{P}^{th} and a Monte Carlo simulation of an ideal machine that samples \mathbf{P}^{th} a finite number of times for our three and (b) four photon cases. The inset histograms show the variation in d expected for a sample size corresponding to the 1421 and 405 counts collected in our three- and four-photon experiments. The distance $d(\mathbf{P}^{\text{exp}}, \mathbf{P}^{\text{th}})$ (red) between our experimental sample and the ideal distribution suggests an underlying systematic inaccuracy. A full model which includes multiphoton emission and photon distinguishability is validated by the distance $d(\mathbf{P}^{\text{exp}}, \mathbf{P}^{\text{mod}})$ (green) which is consistent with statistical variation. (c) The predicted variation in $d(\mathbf{P}^{\text{th}}, \mathbf{P}^{\text{mod}})$ is shown as a function of λ for the photon distinguishabilities V indicated for the three- and (d) four-photon cases. Our experimental conditions are marked (red dot).

all four-fold events.¹ For each experiment, the measured relative frequencies P_S^{exp} for every allowed outcome $|S\rangle$ are shown along with their observed statistical variation. The corresponding theoretical P_S^{th} , calculated using the right-hand side of Eqn. (1), are shown along with their uncertainties arising from the characterization of Λ , described below.

We reconstruct Λ with a series of one- and two-photon transmission measurements to determine its complex-valued elements $\Lambda_{ij} = \tau_{ij}e^{i\phi_{ij}}$ (26). The characterization results for the circuit used in the three-photon experiment are shown in Fig. 2b-c. To obtain the magnitude τ_{ij} , single photons are injected in mode i . The probability of a subsequent detection in mode j is given by $P_1(j, i) = |\Lambda_{ij}|^2 = \tau_{ij}^2$. The phases ϕ_{ij} are determined

¹Eqn. (1) is expected to hold for any $|S\rangle$ and $|T\rangle$, however the classical hardness of sampling $P(S|T)$ is maximized when $S_j, T_i \in \{0, 1\}$ for a given N and Λ .

from two-photon quantum interference measurements. The probability that a photon is detected in each of modes j_1 and j_2 when they are injected in modes i_1 and i_2 is given by $P_2(j_1, j_2, i_1, i_2) = |\Lambda_{i_1j_1}\Lambda_{i_2j_2} + \Lambda_{i_2j_1}\Lambda_{i_1j_2}|^2$. This expression is used to relate the relevant phases ϕ_{ij} given the previously determined magnitudes τ_{ij} . We use a least-squares fit to determine the phases ϕ_{ij} from a complete set of two-photon measurements (23).

To analyze the performance of our QBSM we compare our results to an ideal machine. We quantify the match of two sets of relative frequencies $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ by calculating the L_1 distance $d^{(N)}(\mathbf{P}^{(1)}, \mathbf{P}^{(2)}) = \frac{1}{2} \sum_S |P_S^{(1)} - P_S^{(2)}|$, where N denotes the number of photons in a sample (27). Identical and maximally dissimilar distributions correspond to $d=0$ and $d=1$, respectively. For our experiments we calculate $d^{(N)}(\mathbf{P}^{\text{exp}}, \mathbf{P}^{\text{th}})$ to give $d^{(3)} = 0.094 \pm 0.014$ and $d^{(4)} = 0.097 \pm 0.004$. Even in an ideal QBSM with perfect state prepa-

ration and detection, the statistical variations result in nonzero d . If we substitute for our experimental data a Monte Carlo sampling of \mathbf{P}^{th} with sample size equivalent to our experiments, we instead calculate $d^{(3)}=0.043\pm0.012$ and $d^{(4)}=0.059\pm0.022$. This suggests there are significant contributions to $d(\mathbf{P}^{\text{exp}}, \mathbf{P}^{\text{th}})$ beyond statistical deviation.

Due to experimental limitations, our QBSM occasionally samples distributions other than \mathbf{P}^{th} . The dominant sources of this sampling inaccuracy are multi-photon emission and partial distinguishability of our photon sources. In practice, all single-photon sources produce multiple photons with a finite probability (19). For our PDC sources, the output state is approximately $|00\rangle + \lambda|11\rangle + \lambda^2|22\rangle$, with $\lambda \ll 1$. Both single-photon and undesired multiphoton terms increase with λ . In our three-photon experiments, for example, multiphoton emission from the two PDC sources can lead to input states $|\mathbf{T}\rangle = |021010\rangle$ or $|012020\rangle$, which contribute to three-fold coincident events if photons are lost or emerge in the same output mode. In addition, partial distinguishability of the photons contaminates the distribution sampled by the QBSM by mixing in one- and two-photon interference effects (28). To characterize the photon indistinguishability we performed two-photon Hong-Ou-Mandel experiments, in which interference visibilities were 90% of the expected values on average.

We form a new distribution \mathbf{P}^{mod} that accounts for the effects of multiphoton emission and photon distinguishability (23). The distance $d(\mathbf{P}^{\text{exp}}, \mathbf{P}^{\text{mod}})$, shown by the green point in the insets of Fig. 4a-b, is found to be consistent with the statistical variation due to a finite sample size, for both the three- and four-photon experiments. This suggests we have correctly identified and modeled the significant sources of inaccuracy. Our validated full model may be used to predict the performance of realistic QBSMs and guide the design of future machines with larger N . To investigate how the performance of our QBSM depends on λ and photon distinguishability, we calculate $d(\mathbf{P}^{\text{mod}}, \mathbf{P}^{\text{th}})$ for a range of operating parameters (Fig. 4c-d). In terms of λ , a clear tradeoff is presented between data rate and inaccuracy due to multiphoton emission, which is an intrinsic consequence of using PDC sources. Improvement in photon indistin-

guishability increases the fidelity to the ideal machine, and additionally is thought to enhance the computational power of a QBSM (28).

Our results demonstrate that boson sampling provides a model of quantum-enhanced computation that is experimentally feasible with existing photonic technology. Future generations of QBSMs will benefit from ongoing advances in integrated photonics such as reduced transmission loss, efficient number-resolving detectors (29), and multiplexed (30, 31) or single-emitter (19) photon sources. Such improvements suggest a promising route to large-scale photonic QBSMs, which will provide clear evidence for the computational power of quantum mechanics.

References and Notes

1. D. P. DiVincenzo, *Fortschritte der Physik* **48**, 771 (2000).
2. R. Raussendorf, H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188 (2001).
3. M. A. Nielsen, *Phys. Lett. A* **308**, 96 (2003).
4. A. M. Childs, *Phys. Rev. Lett.* **102**, 180501 (2009).
5. N. B. Lovett, S. Cooper, M. Everitt, M. Trovers, V. Kendon, *Phys. Rev. A* **81** (2010).
6. P. Walther, *et al.*, *Nature* **434**, 169 (2005).
7. C.-Y. Lu, D. E. Browne, T. Yang, J.-W. Pan, *Phys. Rev. Lett.* **99**, 250504 (2007).
8. B. P. Lanyon, *et al.*, *Phys. Rev. Lett.* **99**, 250505 (2007).
9. B. P. Lanyon, *et al.*, *Science* **334**, 57 (2011).
10. E. Lucero, *et al.*, *Nature Phys.* **8**, 719 (2012).
11. E. Knill, R. Laflamme, *Phys. Rev. Lett.* **81**, 5672 (1998).
12. S. P. Jordan, *Quant. Inf. Comput.* **10**, 470 (2010).
13. D. Shepherd, M. J. Bremner, *Proc. R. Soc. A* **465**, 1413 (2009).

14. S. Aaronson, A. Arkhipov, *Proceedings of ACM Symposium on the Theory of Computing, STOC* (2011).
15. E. Caianiello, *Il Nuovo Cimento (1943-1954)* **10**, 1634 (1953).
16. L. Troyansky, N. Tishby, *Proceedings of PhysComp* (1996).
17. L. G. Valiant, *Theoretical Computer Science* **8**, 189 (1979).
18. H. Ryser, *Combinatorial mathematics* (Mathematical Association of America, 1963).
19. M. D. Eisaman, J. Fan, A. Migdall, S. V. Polyakov, *Rev. Sci. Instrum.* **82**, 071101 (2011).
20. B. J. Metcalf, *et al.*, *arXiv:1208.4575v2* (2012).
21. E. Knill, R. Laflamme, G. J. Milburn, *Nature* **409**, 46 (2001).
22. C. K. Hong, Z. Y. Ou, L. Mandel, *Phys. Rev. Lett.* **59**, 2044 (1987).
23. *Materials and methods are available as supplementary material on Science Online .*
24. P. J. Mosley, *et al.*, *Phys. Rev. Lett.* **100**, 133601 (2008).
25. B. J. Smith, D. Kundys, N. Thomas-Peter, P. G. R. Smith, I. A. Walmsley, *Opt. Express* **17**, 264 (2009).
26. A. Laing, J. L. O'Brien, *arXiv:1208.2868v1* (2012).
27. A. Gilchrist, N. K. Langford, M. A. Nielsen, *Phys. Rev. A* **71**, 062310 (2005).
28. P. P. Rohde, *Phys. Rev. A* **86**, 052321 (2012).
29. T. Gerrits, *et al.*, *Phys. Rev. A* **84**, 060301 (2011).
30. A. L. Migdall, D. Branning, S. Castelletto, *Phys. Rev. A* **66**, 053805 (2002).
31. J. Nunn, *et al.*, *arXiv:1208.1534v1* (2012).
32. S. Scheel, *arXiv:0406127v1* (2004).
33. N. Thomas-Peter, *et al.*, *New J. Phys.* **13**, 055024 (2011).
34. M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences)* (Cambridge University Press, 2004).
35. B. J. Smith, P. Mahou, O. Cohen, J. S. Lundeen, I. A. Walmsley, *Opt. Express* **17**, 23589 (2009).

Acknowledgements: We thank Josh Nunn for valuable insights. This work was supported by the EPSRC(EP/C51933/01 and EP/C013840/1), the EC project Q-ESSENCE (248095), the Royal Society, and the AFOSR EOARD. XMJ is supported by EU Marie-Curie Fellowship (PIIF-GA-2011-300820). JS acknowledges support from the United States Air Force Institute of Technology. The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

1 Supplementary Information

In this section we first outline how the boson distribution is given by a set of matrix permanents. We then show how the boson distribution can be accurately predicted by the permanents associated with a non-unitary matrix describing a lossy channel, Λ , if one post-selects on trials where no photons are lost. We describe how one- and two-photon statistics can accurately characterize Λ . Finally, the principal sources of error in this experiment, namely the photon distinguishability and higher order terms from our PDC photon sources, are discussed.

1.1 The boson distribution is given by a set of matrix permanents

We assume N photons are injected into an M mode circuit. We will consider the simplified case where the input (and output) states contain no more than one photon per mode, though the general case is treated elsewhere (32). Without loss of generality, let modes 1 to N contain an input photon, while modes $N + 1$ to M have vacuum inputs. The input state can then be described by

$$|\Psi_{\text{in}}\rangle = |T\rangle = \prod_{i=1}^N \hat{a}_i^\dagger |0\rangle \quad (2)$$

The unitary transformation allows one to evolve the operators according to

$$\hat{a}_i^\dagger = \sum_{j=1}^M U_{ij} \hat{b}_j^\dagger \quad (3)$$

where \hat{a}_i^\dagger and \hat{b}_j^\dagger are creation operators on the i -th input and j -th output mode respectively. We then obtain the output state

$$|\Psi_{\text{out}}\rangle = \prod_{i=1}^N \left(\sum_{j=1}^M U_{ij} \hat{b}_j^\dagger \right) |0\rangle \quad (4)$$

To find the boson distribution, we project our output onto a state $|\mathbf{S}\rangle$ which, in the number state basis, we describe by an N element vector \mathbf{S} , where

S_j gives the mode of the j -th photon. The probability of measuring this state is then

$$P_S = |\langle \mathbf{S} | \psi_{\text{out}} \rangle|^2 \quad (5)$$

$$= \left| \langle 0 | \prod_{i=1}^N \hat{b}_{S_i} \left[\prod_{j=1}^N \left(\sum_{k=1}^M U_{jk} \hat{b}_k^\dagger \right) \right] | 0 \rangle \right|^2$$

The term in square brackets can be expanded and includes M^N terms, as one is selecting N photons from M modes where repetitions are allowed (> 1 boson in a mode). One can rewrite this term in square brackets to give

$$P_S = \left| \langle 0 | \prod_{i=1}^N \hat{b}_{S_i} \left[\sum_{j=1}^{M^N} \left(\prod_{k=1}^N U_{k, \tilde{V}_k^j} \hat{b}_{\tilde{V}_k^j}^\dagger \right) \right] | 0 \rangle \right|^2 \quad (6)$$

where $\tilde{\mathbf{V}}$ is the set of M^N permutations of N photons amongst M modes, repetitions allowed. The tilde notation will be used throughout this paper for a set of permutations. Then, \tilde{V}_k^j indicates the spatial mode of the k -th photon in the j -th permutation. As an example, consider the case with $M = 3$ modes and $N = 2$ input photons, then

$$\begin{aligned} \tilde{V}^1 &= [1, 1] & \tilde{V}^2 &= [1, 2] & \tilde{V}^3 &= [1, 3] \\ \tilde{V}^4 &= [2, 1] & \tilde{V}^5 &= [2, 2] & \tilde{V}^6 &= [2, 3] \\ \tilde{V}^7 &= [3, 1] & \tilde{V}^8 &= [3, 2] & \tilde{V}^9 &= [3, 3] \end{aligned} \quad (7)$$

Let us denote all $N!$ permutations of \mathbf{S} by $\tilde{\mathbf{S}}$ where \tilde{S}_k^j indicates the spatial mode of the k -th photon in the j -th permutation. For example, if we project onto the state $|\mathbf{S}\rangle = |011\rangle$ then $\mathbf{S} = [2, 3]$ and

$$\tilde{S}^1 = [2, 3] \quad \tilde{S}^2 = [3, 2] \quad (8)$$

It is clear we only retain terms from the summation in Eqn. (6) where $\tilde{V}^j \in \tilde{\mathbf{S}}$, otherwise at least one annihilation operator will act on vacuum and give $P_S = 0$. This then leaves us with

$$P_S = \left| \sum_{j=1}^{N!} \prod_{k=1}^N U_{k, \tilde{S}_k^j} \right|^2 \quad (9)$$

The formula for the permanent of an $n \times n$ matrix $A = (a_{ij})$ is

$$\text{Per}(A) = \sum_{i=1}^{n!} \prod_{j=1}^n a_{j, \tilde{\sigma}_j^i} \quad (10)$$

where $\tilde{\sigma}_j^i$ gives the j -th element of the i -th permutation of the numbers $1, 2, \dots, n$. The term inside the modulus of Eqn. (9) has the same form as the matrix permanent in Eqn. (10). Our original unitary, U , can be described by an $M \times M$ matrix. However, it is obvious from Eqn. (9) that, in general, we take the permanent of a subsection of U . Specifically, we only keep rows $1 \rightarrow N$, those rows corresponding to modes with input photons. In addition, we only keep columns corresponding to the elements of S . Let us call this modified subsection of our original unitary $U^{(S,T)}$. Then, using the definition of the permanent we can rewrite Eqn. (9)

$$P(S|T) = \left| \text{Per}(U^{(S,T)}) \right|^2 \quad (11)$$

If one allows the possibility of more than one photon per input and output mode, then a similar analysis yields Eqn. (1) (32).

1.2 Effects of loss

In any experimental implementation of boson sampling there will be losses. These losses, regardless of where they occur, can be modeled as beam splitters that link accessible to inaccessible modes (33). When these losses are considered, it is important to ask whether the boson distribution is still given by a set of matrix permanents and, if so, what linear transformation does that matrix describe.

Let us adopt the convention that modes 1 to M are accessible modes while loss modes are given the labels $M + 1$ to L . There is an $L \times L$ unitary operation describing the evolution of our pure input state, though we must trace over these loss modes at the output, yielding a mixed state over the accessible modes. Let us assume that photons are input into modes 1 to N where $N \leq M$ and we post-select on cases where N photons are detected, by definition, in the accessible modes 1 to M . Then Eqn. (6) becomes

$$P_S = \left| \langle 0 | \prod_{i=1}^N \hat{b}_{S'_i} \left[\sum_{j=1}^{L^N} \left(\prod_{k=1}^N U_{k, \tilde{V}_k^j} \hat{b}_{\tilde{V}_k^j}^\dagger \right) \right] \right|^2 \quad (12)$$

but $S_i \leq M$ as we can only project on accessible modes. Therefore, even though U is an $L \times L$ matrix and the elements of \tilde{V} range from $1 \rightarrow L$,

when we project onto \tilde{S} (all the permutations of S) we are left with

$$P_S = \left| \sum_{j=1}^{N!} \prod_{k=1}^N U_{k, \tilde{S}_k^j} \right|^2 = \left| \text{Per}(U^{(S,T)}) \right|^2 = \left| \text{Per}(\Lambda^{(S,T)}) \right|^2 \quad (13)$$

where $U^{(S,T)}$ is again a modified version of the original unitary but only keeping rows $1 \rightarrow N$ and columns in $S'_i \leq M$. Since these elements always describe the accessible modes, then we can equivalently work in terms of Λ , where $\Lambda_{i,j} = U_{i,j}$ but $i, j \leq M$.

In summary, when post-selecting on no bosons being lost, one can work in terms of Λ , a non-unitary linear transformation that is simply the subsection of U over the accessible modes.

Equivalently, we can describe our system as a noisy (lossy) quantum channel in the operator sum representation (34). This formalism will be useful later when we discuss sources of error from higher order PDC terms. In this picture, the accessible modes are in the system Q , while all inaccessible loss modes form the environment system E . We can describe the transformation induced by our circuit over the full space $Q \otimes E$ by a unitary operation U . Let ρ and σ be the inputs to Q and E respectively, then the output in Q after a projective measurement P_m and tracing over the environment is described by

$$\rho_{\text{out}} = \text{tr}_E(P_m U(\rho \otimes \sigma) U^\dagger P_m) \quad (14)$$

Let the basis for E be described by $|e_k\rangle$ and the initial state of the environment be $\sigma = \sum_j q_j |j\rangle\langle j|$, then we can express Eqn. (14) as

$$\rho_{\text{out}} = \sum_{jk} E_{jk} \rho E_{jk}^\dagger \quad (15)$$

where $E_{jk} = \sqrt{q_j} \langle e_k | P_m U | j \rangle$ are the Kraus operators. We do not directly characterize U , as it extends over the environment which is inaccessible to the experimenter. However, with photons we can assume $\sigma = |0\rangle\langle 0|$, and all of our boson sampling results post-select on the case where no photons are lost to the environment. Therefore, the summation

in Eqn. (15) reduces to only one term with postselection

$$\rho_{\text{out}} = E_{00}\rho E_{00}^\dagger \quad (16)$$

Experimental boson sampling efforts will sample a non-unitary transformation that is equivalent, when post-selecting on no bosons being lost, to the E_{00} Kraus operator.

1.3 Characterizing the unitary over the accessible modes

We want to find $\Lambda_{i,j}$, the elements of the non-unitary linear transformation induced by our lossy circuit, to predict the post-selected boson distribution. Every element is a complex number that can be written $\Lambda_{i,j} = \tau_{i,j}e^{i\phi_{i,j}}$. We can find τ_{i_1,j_1} by coupling light into mode i_1 and monitoring the probability of output in j_1 yielding

$$\begin{aligned} \left| \langle 0 | \hat{b}_{j_1} \hat{a}_{i_1}^\dagger | 0 \rangle \right|^2 &= \left| \langle 0 | \hat{b}_{j_1} \sum_{j=1}^M \Lambda_{i_1 j} \hat{b}_j^\dagger | 0 \rangle \right|^2 \\ &= |\Lambda_{i_1 j_1}|^2 \\ &= |\tau_{i_1 j_1}|^2 \end{aligned} \quad (17)$$

It is sufficient to measure the ratio of values between output ports, $(\tau_{i_1,j_1}/\tau_{i_1,j_2})^2$, as this describes the transformation induced by our circuit up to a constant factor for each input, x_i . If Λ is the true transformation, then this procedure gives us $X\Lambda$ where X is a diagonal matrix with entries $X_{ii} = x_i$. However, as $\text{Per}(X\Lambda) = (\prod_i x_i)\text{Per}(\Lambda)$ then Eqn. (1) shows, since we always launch the same input state, every portion of our boson distribution is multiplied by a constant factor that cancels in the normalization. These data are collected by periodically pausing boson sampling data collection, blocking two of the three photon inputs, and measuring the relative power in each output mode. We thus obtain accurate values for τ as well as a variance that is important for calculating the error bars in Fig. 3.

One can then use two photon interference to find $\phi_{i,j}$ (26). If one inputs two photons into modes i_1, i_2 and detect in modes j_1, j_2 , then we have $|S\rangle = \hat{a}_{j_1}^\dagger \hat{a}_{j_2}^\dagger |0\rangle$ and $|T\rangle = \hat{a}_{i_1}^\dagger \hat{a}_{i_2}^\dagger |0\rangle$. If single,

indistinguishable photons are used, then the probability of post-selecting this output is

$$\begin{aligned} P_{\text{indist}} &= \left| \text{Per}(\Lambda^{(S,T)}) \right|^2 \\ &= \left| \Lambda_{11}^{(S,T)} \Lambda_{22}^{(S,T)} + \Lambda_{12}^{(S,T)} \Lambda_{21}^{(S,T)} \right|^2 \\ &= (\tau_{i_1,j_1} \tau_{i_2,j_2})^2 + (\tau_{i_1,j_2} \tau_{i_2,j_1})^2 \\ &\quad + 2\tau_{i_1,j_1} \tau_{i_1,j_2} \tau_{i_2,j_1} \tau_{i_2,j_2} \\ &\quad \times \cos(\phi_{i_1,j_1} - \phi_{i_1,j_2} - \phi_{i_2,j_1} + \phi_{i_2,j_2}) \end{aligned} \quad (18)$$

If the photons launched into the individual modes are distinguishable, then we get the incoherent sum of their individual statistics. We again take a matrix permanent, but as this is an incoherent process one finds

$$P_{\text{dist}} = (\tau_{i_1,j_1} \tau_{i_2,j_2})^2 + (\tau_{i_1,j_2} \tau_{i_2,j_1})^2 \quad (19)$$

One can then perform a Hong-Ou-Mandel interference experiment and find that the resulting dip visibility is

$$\begin{aligned} V &= \frac{P_{\text{indist}} - P_{\text{dist}}}{P_{\text{indist}}} \\ &= \frac{2\tau_{i_1,j_1} \tau_{i_1,j_2} \tau_{i_2,j_1} \tau_{i_2,j_2}}{(\tau_{i_1,j_1} \tau_{i_2,j_2})^2 + (\tau_{i_1,j_2} \tau_{i_2,j_1})^2} \end{aligned} \quad (20)$$

$$\times \cos(\phi_{i_1,j_1} + \phi_{i_2,j_2} - \phi_{i_1,j_2} - \phi_{i_2,j_1}) \quad (21)$$

By measuring the dip visibility, and using the known $\tau_{i,j}$ one can find $|\phi_{i_1,j_1} + \phi_{i_2,j_2} - \phi_{i_1,j_2} - \phi_{i_2,j_1}|$. Repeating this procedure for all accessible two photon dips, and applying additional constraints one can determine $\phi_{i,j}$ (26). In theory, one can directly use Eqn. (18) to constrain the phases similarly to Eqn. (21). However, the former relation is useful only if one post-selects on the entire N photon boson distribution which, in our case, requires number-resolving detectors. For our circuit topology, we encounter an overconstrained problem as we measure more interference visibilities than unknown ϕ . Therefore, we run a least squares minimization to find the set of ϕ that best fits our two photon interference data. This characterization process is efficient, as a general linear transformation over M modes can be described by $\mathcal{O}(M^2)$ parameters, requiring $\mathcal{O}(M^2)$ measurements with this technique.

With $\Lambda_{i,j}$ experimentally determined over the accessible modes, one can predict the post-selected boson distribution for any input/output $|S\rangle, |T\rangle$ by constructing $\Lambda^{(S,T)}$ from Λ and taking the permanent according to Eqn. (13). Thus, we effectively use the one and two photon boson distributions to characterize our non-unitary operation over the accessible modes. One can then take various matrix permanents of this non-unitary matrix to predict the boson distribution for *any* N .

1.4 Sources of error

The computational difficulty for a classical machine to even verify the accuracy of QBSM output scales exponentially in N , making it vital to check the accuracy of these devices at low N where classical computer validation remains easy. In this paper, we have benchmarked the accuracy of our QBSMs by inferring the probability distribution from our data, labeled \mathbf{P}^{exp} , and comparing it to the distribution obtained from Equation (1), labeled \mathbf{P}^{th} . Throughout the text, we quantify the distance between two probability distributions via the L_1 distance, $d(\mathbf{P}^{(1)}, \mathbf{P}^{(2)}) = \frac{1}{2} \sum_i |\mathbf{P}_i^{(1)} - \mathbf{P}_i^{(2)}|$.

Our method of benchmarking QBSM accuracy will always yield a nonzero d due to the finite number of samples collected. We perform a Monte Carlo simulation of a QBSM that perfectly samples \mathbf{P}^{th} as a function of the number of counts collected, shown in Figure 4(a)-(b), to show the rate at which d asymptotically approaches zero as the number of samples collected increases. In the inset histograms, we show the range of outputs for this ideal QBSM for the actual number of experimental counts collected in the three and four photon cases, while the red dots indicate $d(\mathbf{P}^{\text{th}}, \mathbf{P}^{\text{exp}})$. These two probability distributions show close agreement in Figure 3, however a comparison of the histogram and red dots in Figure 4(a)-(b) indicates there is an additional source of error beyond the finite number of samples.

Due to experimental limitations, occasionally we see sample distributions other than \mathbf{P}^{th} . We model the effect of two such imperfections, photon distinguishability and higher order terms from our PDC sources, and form a new distribution \mathbf{P}^{mod} that accounts for these effects. We find that the

new $d(\mathbf{P}^{\text{exp}}, \mathbf{P}^{\text{mod}})$, indicated with the green dot in 4(a)-(b), comes well within the output variance of an ideal machine. This indicates we have correctly diagnosed and modeled the principal sources of experimental error, which will be important in guiding designs of future, larger N , QBSMs.

1.4.1 Effect of using heralded single photon sources

The parametric downconversion sources we use to generate our photons actually generate a two-mode squeezed state that is given by

$$|\Psi_{\text{PDC}}\rangle = \sqrt{1 - \lambda^2} \sum_{n=0}^{\infty} \lambda^n |nn\rangle \quad (22)$$

where $0 \leq \lambda < 1$ is the squeezing parameter whose magnitude is determined, in part, by the type of crystal and pump power used. For our experiment we wish to minimize higher order terms (ex. $\lambda^2|22\rangle$ for the three photon QBSM) and so we deliberately lower our pump power as much as possible while maintaining a feasible count rate. For the three-photon experiment, $\lambda = \sqrt{0.011}$ and for the four-photon experiment $\lambda = \sqrt{0.023}$. However, even in this case we will sometimes inject more than N photons into our circuit which, due to losses, could be observed as an N -fold detection at the output.

Due to our characterization method, we have no information about what such terms will be. In the operator sum representation, our characterization only determines the E_{00} Kraus operator, which describes the transformation of our input state when the environment, which includes all loss modes, starts and ends with zero photons. For example, if we instead inject five photons and lose two, then this process is described by the E_{20} operator, about which we have no information.

To model the effect of using such squeezed sources, we start with a circuit with the same topology used in the experiment. Non-uniform losses throughout the circuit can then be modeled by adding beam splitters that link the depicted accessible modes shown in the figure to inaccessible loss modes, where the beam splitter reflectivity indicates the loss in that channel (33). Such ‘loss beam splitters’ are added throughout the circuit. We cannot directly characterize these losses, however we

can estimate them numerically. The characterized linear transformation \mathbf{A} is a function of these losses as well as the fabricated interferometers shown in Fig 2. We have performed a loss-independent characterization of the interferometers (20), and then input this data into a genetic algorithm to find the relative losses that best reproduces the characterized \mathbf{A} .

With knowledge of these losses, we reproduce an actual unitary linear transformation \mathbf{U} that extends over both the $M = 6$ modes as well as all loss modes and can be used to accurately predict the effect of higher order PDC terms. We use Eqn. (1) to find the probability distributions when the input is $|\mathbf{T}\rangle = |011010\rangle, |022010\rangle$ or $|011020\rangle$ which we label \mathbf{P}^{111} , \mathbf{P}^{221} and \mathbf{P}^{112} respectively. For example, the probability of obtaining $|\mathbf{S}\rangle = |111000\rangle$ given input $|\mathbf{T}\rangle = |022010\rangle$ is the summation of all terms where three photons appear in the first three accessible modes and two photons appear in any combination of loss modes. The higher order probability distributions, \mathbf{P}^{221} and \mathbf{P}^{112} , are weighted by λ^2 which is obtained via a conditional second order correlation measurement, $g^{(2)}(0)$ (35). As λ^2 is small, we only consider the first higher order terms from each source.

1.4.2 Photon Distinguishability

Boson sampling assumes indistinguishable bosons, while experimental implementations will always have some distinguishability. Assuming pure inputs we follow the notation of (28) to write our input state as

$$|\psi_{\text{in}}\rangle = \prod_i^N \left(\alpha A_{\xi_0,i}^\dagger + \sqrt{1 - \alpha^2} A_{\xi_i,i}^\dagger \right) |0\rangle \quad (23)$$

where N is again the number of photons, α is a distinguishability parameter, and $A_{\xi_j,i}^\dagger$ is the creation operator for photon i in mode ξ_j . By analyzing the reduction in HOM dip visibility at a beamsplitter inside our circuit we estimate $\alpha = 0.974$ in our experiment.

If one photon is distinguishable from the others, then the new probability distribution is given by the permanents of $N - 1$ matrices which are incoherently summed. For example, assume an input state

$|\mathbf{T}\rangle = |1\rangle^\tau |11000\rangle$ where τ labels a distinguishable photon, then for a unitary transformation U the probability of obtaining an output $|\mathbf{S}\rangle$ is

$$\begin{aligned} P(\mathbf{S}|\mathbf{T}) = & \left| U_{11}^{(\mathbf{S},\mathbf{T})} \right|^2 \left| U_{22}^{(\mathbf{S},\mathbf{T})} U_{33}^{(\mathbf{S},\mathbf{T})} + U_{23}^{(\mathbf{S},\mathbf{T})} U_{32}^{(\mathbf{S},\mathbf{T})} \right|^2 \\ & + \left| U_{12}^{(\mathbf{S},\mathbf{T})} \right|^2 \left| U_{21}^{(\mathbf{S},\mathbf{T})} U_{33}^{(\mathbf{S},\mathbf{T})} + U_{23}^{(\mathbf{S},\mathbf{T})} U_{31}^{(\mathbf{S},\mathbf{T})} \right|^2 \\ & + \left| U_{13}^{(\mathbf{S},\mathbf{T})} \right|^2 \left| U_{21}^{(\mathbf{S},\mathbf{T})} U_{32}^{(\mathbf{S},\mathbf{T})} + U_{22}^{(\mathbf{S},\mathbf{T})} U_{31}^{(\mathbf{S},\mathbf{T})} \right|^2 \end{aligned} \quad (24)$$

where the terms in parentheses are permanents of 2×2 matrices. We calculate these probability distributions when one photon is distinguishable and weight them by $|\alpha^2 \sqrt{1 - \alpha^2}|^2$ and $|\alpha^3 \sqrt{1 - \alpha^2}|^2$, the probability that one photon is distinguishable from the others for the two and three photon cases respectively. As α is large, we ignore the case when two photons are distinguishable.